10/538110
PCT/IB 03/05766
05.12.03

Europäisches Patentamt | European Patent Office | Office européen des brevets

REC'D 22 DEC 2003
WIPO PCT

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

02293062.2

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

BEST AVAILABLE COPY

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02293062.2
Demande no:

Anmeldetag:
Date of filing: 11.12.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Video encoding method

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/26

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SI SK

"VIDEO ENCODING METHOD"

FIELD OF THE INVENTION

The present invention generally relates to the field of data compression and, more specifically, to a method of encoding a sequence of frames, composed of picture elements (pixels), by means of a three-dimensional (3D) subband decomposition involving a filtering step applied, in the sequence considered as a 3D volume, to the spatial-temporal data which correspond in said sequence to each one of successive groups of frames (GOFs), these GOFs being themselves subdivided into successive pairs of frames (POFs) including a so-called previous frame and a so-called current frame, said decomposition being applied to said GOFs together with motion estimation and compensation steps performed in each GOF on saids POFs and on corresponding pairs of low-frequency temporal subbands (POSs) obtained at each temporal decomposition level.

The invention also relates to a computer programme comprising a set of instructions for the implementation of said encoding method, when said programme is carried out by a processor included in an encoding device.

BACKGROUND OF THE INVENTION

In recent years, three-dimensional (3D) subband analysis, based on a 3D, or (2D+t), wavelet decomposition of a sequence of frames considered as a 3D volum has been more and more studied for video compression. The coefficients generated by the wavelet transform constitute a hierarchical pyramid in which the spatio-temporal relationship is defined thanks to 3D orientation trees evidencing the parent-offspring dependencies between coefficients, and the in-depth scanning of the generated coefficients in the hierarchical trees and a progressive bitplane encoding technique lead to a desired quality scalability. The practical stage for this approach is to generate motion compensated temporal subbands using a simple two taps wavelet filter, as illustrated in Fig.1 for a group of frames (GOF) of eight frames.

In the illustrated implementation, the input video sequence is divided into Groups of Frames (GOFs), and each GOF, itself subdivided into successive couples of frames (that are as many inputs for a so-called Motion-Compensated Temporal Filtering, or MCTF module), is first motion-compensated (MC) and then temporally filtered (TF). The resulting low frequency (L) temporal subbands of the first temporal decomposition level are further filtered (TF), and the process may stop after an arbitrary number of decompositions resulting in one or more low frequency subbands called root temporal subbands (in the illustration, an example with two decomposition levels resulting in two root subbands is presented). In the example of Fig.1, the frames of the illustrated group are referenced F1 to F8, and the dotted arrows correspond to a high-pass temporal

filtering, while the other ones correspond to a low-pass temporal filtering. Two stages of decomposition are shown (L and H = first stage ; LL and LH = second stage). At each temporal decomposition level of the illustrated group of 8 frames, a group of motion vector fields is generated (in the present example, MV4 at the first level, MV3 at the second one).

When a Haar multiresolution analysis is used for the temporal decomposition, since one motion vector field is generated between every two frames in the considered group of frames at each temporal decomposition level, the number of motion vector fields is equal to half the number of frames in the temporal subband, i.e. four at the first level of motion vector fields and two at the second one. Motion estimation (ME) and motion compensation (MC) are only performed every two frames of the input sequence (generally in the forward way), due to the temporal down-sampling by two of the simple wavelet filter. Using these very simple filters, each low frequency temporal subband (L) represents a temporal average of the input couples of frames, whereas the high frequency one (H) contains the residual error after the MCTF step.

Unfortunately, the motion compensated temporal filtering may raise the problem of unconnected picture elements (or pixels), which are not filtered at all (or also the problem of double-connected pixels, which are filtered twice). The number of unconnected pixels represents a weakness of a 3D subband codec approaches because it highly impacts the resulting picture quality (particularly in occlusion regions). It is especially true for high motion sequences or for final temporal decomposition levels, where the temporal correlation is not good. The number of these unconnected pixels depends on the dense motion vector field that has been generated by the motion estimation.

Current criteria for optimal motion vector search used in motion estimators do not take into account the number of unconnected pixels that will be the result of motion compensation. Most sophisticated algorithms use a rate/distortion criterion which tends to minimize a cost function that depends on the displaced difference energy (distortion) and the number of bits spent to transmit the motion vector (rate). For example, the motion search returns the motion vector that minimises :

$$J(\mathbf{m}) = SAD(s, c(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m} - \mathbf{p}) \qquad (1)$$

with $\mathbf{m} = (m_x, m_y)^T$ being the motion vector, $\mathbf{p} = (p_x, p_y)^T$ being the prediction for the motion vector, and λ_{MOTION} being the Lagrange multiplier. The rate term $R(\mathbf{m} - \mathbf{p})$

represents the motion information only and SAD is used as distortion measure. It is computed as :

$$SAD(s, c(\mathbf{m})) = \sum_{x=1,y=1}^{B,B} \left| s[x,y] - c[x - m_x, y - m_y] \right| \quad (2)$$

with s being the original video signal, c being the coded video signal and B being the block size (note that B can be 1). Unfortunately, these algorithms do not take into account the distortion introduced by unconnected pixels during the inverse motion compensation because usually these optimizations are applied to hybrid coding for which the inverse motion compensation is not performed.

SUMMARY OF THE INVENTION

It is therefore an object of the invention to avoid such a drawback and to propose a video encoding method in which the set of unconnected pixels is taken into account in the distortion measure.

To this end, the invention relates to a method of encoding a sequence of frames, composed of picture elements (pixels), by means of a three-dimensional (3D) subband decomposition involving a filtering step applied, in the sequence considered as a 3D volume, to the spatial-temporal data which correspond in said sequence to each one of successive groups of frames (GOFs), these GOFs being themselves subdivided into successive pairs of frames (POFs) including a so-called previous frame and a so-called current frame, said decomposition being applied to said GOFs together with motion estimation and compensation steps performed in each GOF on saids POFs and on corresponding pairs of low-frequency temporal subbands (POSs) obtained at each temporal decomposition level, this process of motion compensated temporal filtering leading in the previous frames on the one hand to connected pixels, that are filtered along a motion trajectory corresponding to motion vectors defined by means of said motion estimation steps, and on the other hand to a residual number of so-called unconnected pixels, that are not filtered at all, each motion estimation step comprising a motion search provided for returning a motion vector that minimizes a cost function depending at least on a distorsion criterion involving a distortion measure, said measure distorsion being also applied to the set of said unconnected pixels.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described, by way of example, with reference to the accompanying drawing in which :

- Fig.1 shows a temporal multiresolution analysis with motion compensation.

DETAILED DESCRIPTION OF THE INVENTION

Because unconnected pixels highly participate to the quality degradation of the inverse motion compensated image, the set of unconnected pixels is, according to the invention, taken into account in the distortion measure. To this end, it is here proposed to introduce a new rate/distortion criterion that extends equation taking into account the unconnected pixels phenomenon. This is illustrated in equations (3) and (4) :

$$K(\mathbf{m}) = J(\mathbf{m}) + \lambda_{UNCONNECTED} \cdot D(S_{UNCONNECTED}(\mathbf{m})) \qquad (3)$$

$$K(\mathbf{m}) = SAD(s, c(\mathbf{m})) + \lambda_{UNCONNECTED} \cdot D(S_{UNCONNECTED}(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m} - \mathbf{p}) \qquad (4)$$

with $D(S_{UNCONNECTED}(\mathbf{m}))$ being the distortion measure for the set $S_{UNCONNECTED}$ of unconnected pixels resulting from motion vector $\mathbf{m}$. Several distortion measures can be applied to the set of unconnected pixels. A very simple measure is preferably the count of unconnected pixels for the motion vector under study.

It can be noted that the real set of unconnected pixels resulting from a motion search can be computed only when the motion vectors information is available for the whole frame. Therefore, an optimal solution can hardly be achievable (in fact a complex set of minimisation criteria for the whole frame should be solved), and a sub-optimal implementation is therefore proposed.

This implementation is not recursive and can be considered as a simple way to take into account the distortion due to unconnected pixels. For a given part of the image to be motion compensated (a part of the image can be a pixel, a block of pixels , a macroblock of pixels or any region provided that the set of parts covers the whole image without any overlapping) and for a given motion vector candidate $\mathbf{m}$, a temporary inverse motion compensation is applied, the set of unconnected pixels is identified and $D(S_{UNCONNECTED}(\mathbf{m}))$ can be evaluated. The current $K(\mathbf{m})$ value can be computed and compared to the current minimum value $K_{min}(\mathbf{m})$ to check if the candidate motion vector brings a lower $K(\mathbf{m})$ value. When all the candidate have been tested, the (final) inverse motion compensation is applied to the best candidate (identifying connected and unconnected pixels). The next part of the image can then be processed, and so on up to a complete processing of the whole image.

CLAIMS :

1. A method of encoding a sequence of frames, composed of picture elements (pixels), by means of a three-dimensional (3D) subband decomposition involving a filtering step applied, in the sequence considered as a 3D volume, to the spatial-temporal data which correspond in said sequence to each one of successive groups of frames (GOFs), these GOFs being themselves subdivided into successive pairs of frames (POFs) including a so-called previous frame and a so-called current frame, said decomposition being applied to said GOFs together with motion estimation and compensation steps performed in each GOF on saids POFs and on corresponding pairs of low-frequency temporal subbands (POSs) obtained at each temporal decomposition level, this process of motion compensated temporal filtering leading in the previous frames on the one hand to connected pixels, that are filtered along a motion trajectory corresponding to motion vectors defined by means of said motion estimation steps, and on the other hand to a residual number of so-called unconnected pixels, that are not filtered at all, each motion estimation step comprising a motion search provided for returning a motion vector that minimizes a cost function depending at least on a distorsion criterion involving a distortion measure, said measure distorsion being also applied to the set of said unconnected pixels.

2. An encoding method according to claim 1, in which said motion search is provided for minimizing the following expression (1) :

$$J(\mathbf{m}) = SAD(s, c(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m} - \mathbf{p}) \qquad (1)$$

with $\mathbf{m} = (m_x, m_y)^T$ being the motion vector, $\mathbf{p} = (p_x, p_y)^T$ being the prediction for the motion vector, λ_{MOTION} being the Lagrange multiplier, the rate term $R(\mathbf{m}-\mathbf{p})$ representing the motion information only, and SAD used as distortion measure being computed as :

$$SAD(s, c(\mathbf{m})) = \sum_{x=1, y=1}^{B,B} \left| s[x, y] - c[x - m_x, y - m_y] \right| \qquad (2)$$

with s being the original video signal, c being the coded video signal and B being the block size, and in which the distorsion criterion extends equation (1), taking into account the unconnected pixels phenomenon for the minimizing operation that is now applied to the following expression (3) :

$$K(\mathbf{m}) = J(\mathbf{m}) + \lambda_{UNCONNECTED} \cdot D(S_{UNCONNECTED}(\mathbf{m})) \qquad (3$$

or $$K(\mathbf{m}) = SAD(s, c(\mathbf{m})) + \lambda_{UNCONNECTED} \cdot D(S_{UNCONNECTED}(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m} - \mathbf{p}) \qquad (4)$$

with $D(S_{UNCONNECTED}(\mathbf{m}))$ being the distortion measure for the set $S_{UNCONNECTED}$ of unconnected pixels resulting from the motion vector $\mathbf{m}$.

3. An encoding method according to claim 2, said method including, for taking into account the distortion due to the unconnected pixels, the following steps, successively applied to each part of the whole image to be motion-compensated :

(a) for the considered part of the image and for a given motion vector candidate m, a temporary inverse motion compensation is applied ;

(b) the set of unconnected pixels is identified ;

(c) $D(S_{UNCONNECTED}(m))$ is evaluated ;

(d) the current K(m) value is computed and compared to the current minimum value $K_{min}(m)$ to check if the motion vector candidate brings a lower K(m) value ;

(e) when all the candidates have been tested, a final inverse motion compensation is applied to the best candidate ;

(f) the steps (a) to (e) are then applied to the next part of the image that can be similarly processed, said part of the image being a pixel, a block of pixels, a macroblock of pixels or any region provided that the set of parts covers the whole image without any overlapping.

Abstract

The invention relates to a method of encoding a sequence of frames, composed of picture elements (pixels), by means of a three-dimensional (3D) subband decomposition involving a filtering step applied, in the sequence considered as a 3D volume, to the spatial-temporal data which correspond in said sequence to each one of successive groups of frames (GOFs), and to a non-recursive implementation of said method. The GOFs are themselves subdivided into successive pairs of frames (POFs) including a so-called previous frame and a so-called current frame, and the decomposition is applied to said GOFs together with motion estimation and compensation steps performed in each GOF on saids POFs and on corresponding pairs of low-frequency temporal subbands (POSs) obtained at each temporal decomposition level. The process of motion compensated temporal filtering leading in the previous frames on the one hand to connected pixels, that are filtered along a motion trajectory corresponding to motion vectors defined by means of said motion estimation steps, and on the other hand to a residual number of so-called unconnected pixels, that are not filtered at all, each motion estimation step comprises a motion search provided for returning a motion vector that minimizes a cost function depending at least on a distorsion criterion, said criterion taking into account the unconnected pixels phenomenon for the minimizing operation.

Fig.1

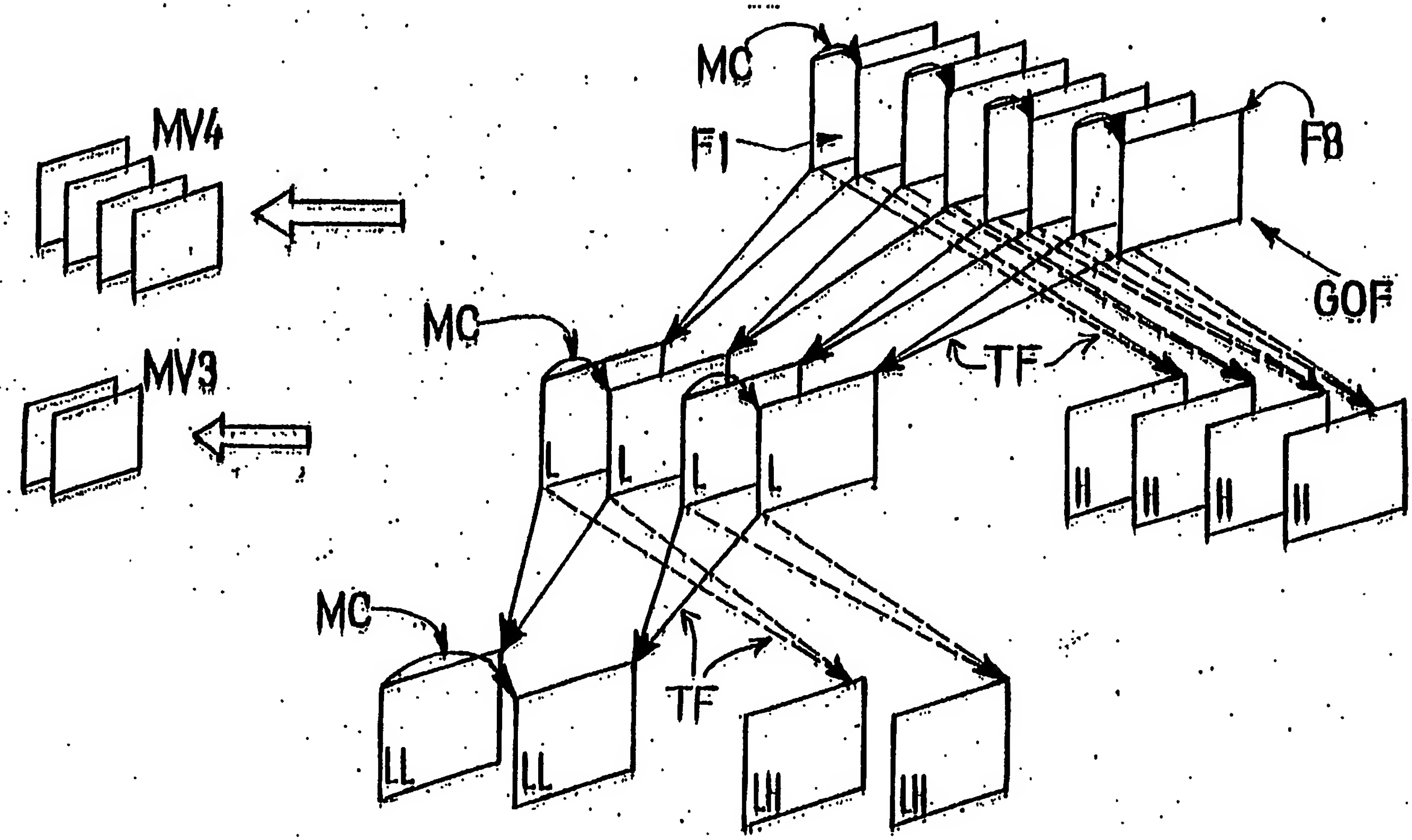

FIG.1

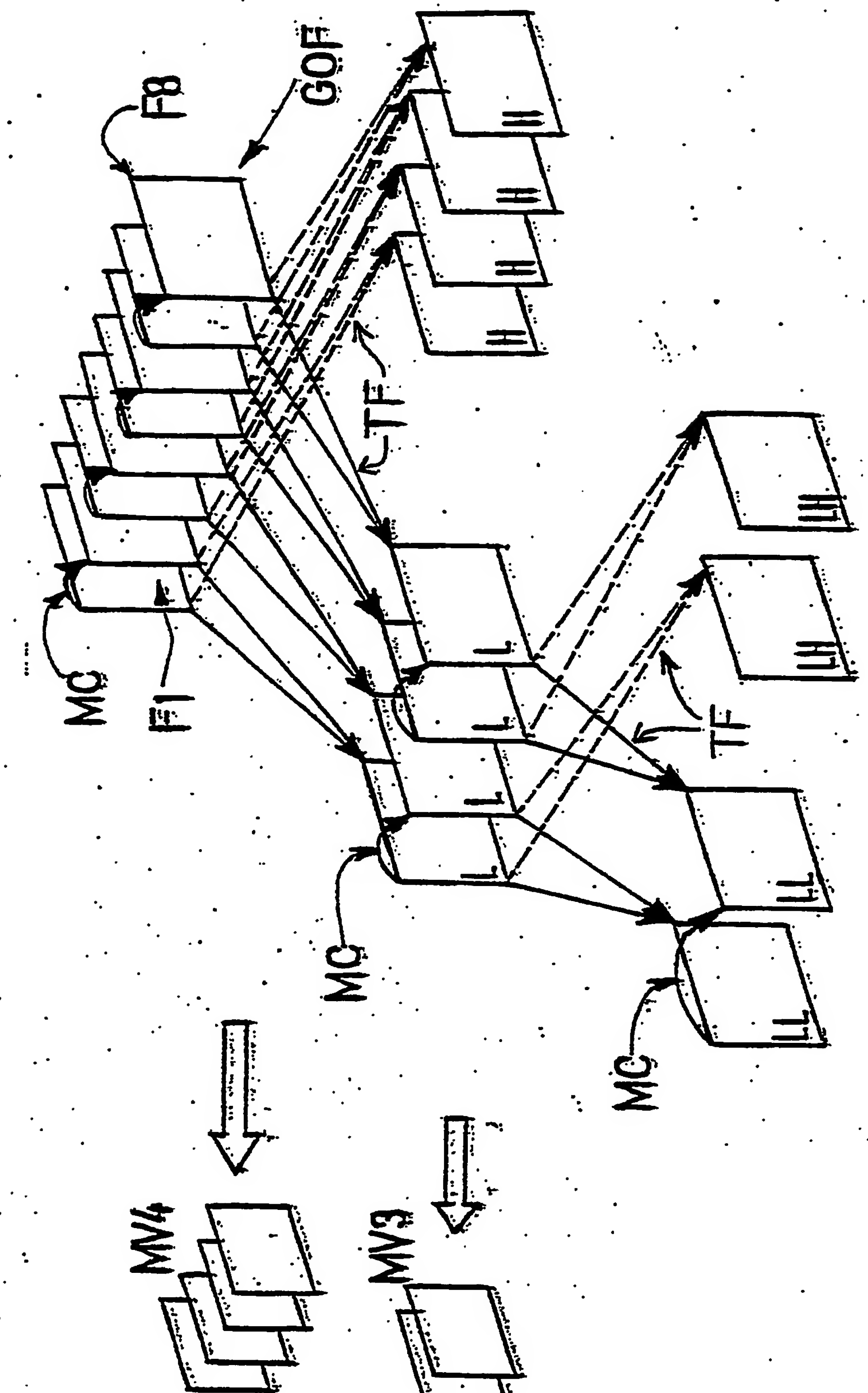

FIG.1